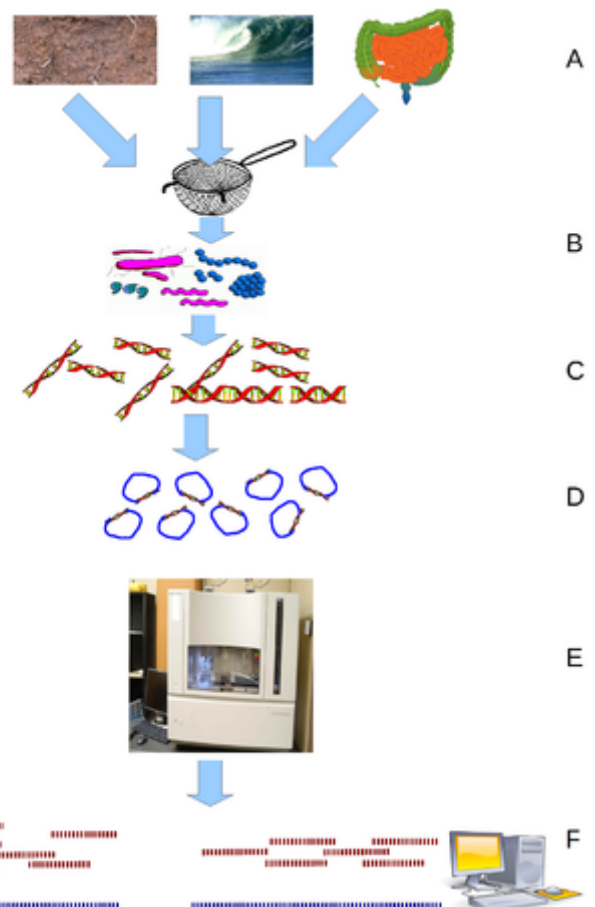


# Metagenomics

**Metagenomics** is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as **environmental genomics**, **ecogenomics** or **community genomics**.

While traditional microbiology and microbial genome sequencing and genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods.<sup>[2]</sup>

Because of its ability to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world.<sup>[3]</sup> As the price of DNA sequencing continues to fall, metagenomics now allows microbial ecology to be investigated at a much greater scale and detail than before. Recent studies use either "shotgun" or PCR directed sequencing to get largely unbiased samples of all genes from all the members of the sampled communities.<sup>[4]</sup>



In metagenomics, the genetic materials (DNA, **C**) are extracted directly from samples taken from the environment (e.g. soil, sea water, human gut, **A**) after filtering (**B**), and are sequenced (**E**) after multiplication by cloning (**D**) in an approach called shotgun sequencing. These short sequences can then be put together again using assembly methods (**F**) to deduce the individual genomes or parts of genomes that constitute the original environmental sample. This information can then be used to study the species diversity and functional potential of the microbial community of the environment.<sup>[1]</sup>

## Contents

### Etymology

### History

### Sequencing

Shotgun metagenomics

High-throughput sequencing

### Bioinformatics

Sequence pre-filtering

Assembly

Gene prediction

Species diversity

Data integration

Comparative metagenomics

**Data analysis**

Community metabolism

Metatranscriptomics

Viruses

**Applications**

Agriculture

Biofuel

Biotechnology

Ecology

Environmental remediation

Gut microbe characterization

Infectious disease diagnosis

**See also**

**References**

**External links**

## Etymology

---

The term "metagenomics" was first used by Jo Handelsman, Jon Clardy, Robert M. Goodman, Sean F. Brady, and others, and first appeared in publication in 1998.<sup>[5]</sup> The term metagenome referenced the idea that a collection of genes sequenced from the environment could be analyzed in a way analogous to the study of a single genome. In 2005, Kevin Chen and Lior Pachter (researchers at the University of California, Berkeley) defined metagenomics as "the application of modern genomics technique without the need for isolation and lab cultivation of individual species".<sup>[6]</sup>

## History

---

Conventional sequencing begins with a culture of identical cells as a source of DNA. However, early metagenomic studies revealed that there are probably large groups of microorganisms in many environments that cannot be cultured and thus cannot be sequenced. These early studies focused on 16S ribosomal RNA (rRNA) sequences which are relatively short, often conserved within a species, and generally different between species. Many 16S rRNA sequences have been found which do not belong to any known cultured species, indicating that there are numerous non-isolated organisms. These surveys of ribosomal RNA genes taken directly from the environment revealed that cultivation based methods find less than 1% of the bacterial and archaeal species in a sample.<sup>[2]</sup> Much of the interest in metagenomics comes from these discoveries that showed that the vast majority of microorganisms had previously gone unnoticed.

Early molecular work in the field was conducted by Norman R. Pace and colleagues, who used PCR to explore the diversity of ribosomal RNA sequences.<sup>[7]</sup> The insights gained from these breakthrough studies led Pace to propose the idea of cloning DNA directly from environmental samples as early as 1985.<sup>[8]</sup> This led to the first report of isolating and cloning bulk DNA from an environmental sample, published by Pace and colleagues in 1991<sup>[9]</sup> while Pace was in the Department of Biology at Indiana University. Considerable efforts ensured that these were not PCR false positives and supported the existence of a complex community of unexplored species. Although this methodology was limited to exploring highly conserved, non-protein coding genes, it did support early microbial morphology-based observations that

diversity was far more complex than was known by culturing methods. Soon after that, Healy reported the metagenomic isolation of functional genes from "zoolibraries" constructed from a complex culture of environmental organisms grown in the laboratory on dried grasses in 1995.<sup>[10]</sup> After leaving the Pace laboratory, Edward DeLong continued in the field and has published work that has largely laid the groundwork for environmental phylogenies based on signature 16S sequences, beginning with his group's construction of libraries from marine samples.<sup>[11]</sup>

In 2002, Mya Breitbart, Forest Rohwer, and colleagues used environmental shotgun sequencing (see below) to show that 200 liters of seawater contains over 5000 different viruses.<sup>[12]</sup> Subsequent studies showed that there are more than a thousand viral species in human stool and possibly a million different viruses per kilogram of marine sediment, including many bacteriophages. Essentially all of the viruses in these studies were new species. In 2004, Gene Tyson, Jill Banfield, and colleagues at the University of California, Berkeley and the Joint Genome Institute sequenced DNA extracted from an acid mine drainage system.<sup>[13]</sup> This effort resulted in the complete, or nearly complete, genomes for a handful of bacteria and archaea that had previously resisted attempts to culture them.<sup>[14]</sup>

Beginning in 2003, Craig Venter, leader of the privately funded parallel of the Human Genome Project, has led the Global Ocean Sampling Expedition (GOS), circumnavigating the globe and collecting metagenomic samples throughout the journey. All of these samples are sequenced using shotgun sequencing, in hopes that new genomes (and therefore new organisms) would be identified. The pilot project, conducted in the Sargasso Sea, found DNA from nearly 2000 different species, including 148 types of bacteria never before seen.<sup>[15]</sup> Venter has circumnavigated the globe and thoroughly explored the West Coast of the United States, and completed a two-year expedition to explore the Baltic, Mediterranean and Black Seas. Analysis of the metagenomic data collected during this journey revealed two groups of organisms, one composed of taxa adapted to environmental conditions of 'feast or famine', and a second composed of relatively fewer but more abundantly and widely distributed taxa primarily composed of plankton.<sup>[16]</sup>

In 2005 Stephan C. Schuster at Penn State University and colleagues published the first sequences of an environmental sample generated with high-throughput sequencing, in this case massively parallel pyrosequencing developed by 454 Life Sciences.<sup>[17]</sup> Another early paper in this area appeared in 2006 by Robert Edwards, Forest Rohwer, and colleagues at San Diego State University.<sup>[18]</sup>

## Sequencing

---

Recovery of DNA sequences longer than a few thousand base pairs from environmental samples was very difficult until recent advances in molecular biological techniques allowed the construction of libraries in bacterial artificial chromosomes (BACs), which provided better vectors for molecular cloning.<sup>[20]</sup>

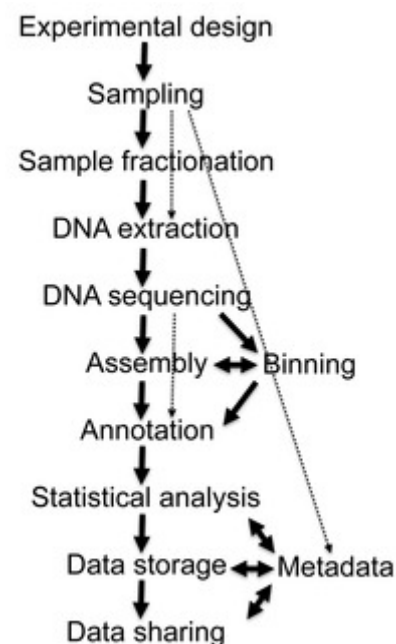
## Shotgun metagenomics

Advances in bioinformatics, refinements of DNA amplification, and the proliferation of computational power have greatly aided the analysis of DNA sequences recovered from environmental samples, allowing the adaptation of shotgun sequencing to metagenomic samples (known also as whole metagenome shotgun or WMGS sequencing). The approach, used to sequence many cultured microorganisms and the human genome, randomly shears DNA, sequences many short sequences, and reconstructs them into a consensus sequence. Shotgun sequencing reveals genes present in environmental samples. Historically, clone libraries were used to facilitate this sequencing. However, with advances in high throughput sequencing technologies, the cloning step is no longer necessary and greater yields of sequencing data can be obtained without this labour-intensive bottleneck step. Shotgun metagenomics provides information both about which organisms are present and what metabolic processes are possible in the community.<sup>[21]</sup> Because the collection of DNA from an environment is largely uncontrolled, the most abundant organisms in an

environmental sample are most highly represented in the resulting sequence data. To achieve the high coverage needed to fully resolve the genomes of under-represented community members, large samples, often prohibitively so, are needed. On the other hand, the random nature of shotgun sequencing ensures that many of these organisms, which would otherwise go unnoticed using traditional culturing techniques, will be represented by at least some small sequence segments.<sup>[13]</sup>

## High-throughput sequencing

An advantage to high throughput sequencing is that this technique does not require cloning the DNA before sequencing, removing one of the main biases and bottlenecks in environmental sampling. The first metagenomic studies conducted using high-throughput sequencing used massively parallel 454 pyrosequencing.<sup>[17]</sup> Three other technologies commonly applied to environmental sampling are the Ion Torrent Personal Genome Machine, the Illumina MiSeq or HiSeq and the Applied Biosystems SOLiD system.<sup>[22]</sup> These techniques for sequencing DNA generate shorter fragments than Sanger sequencing; Ion Torrent PGM System and 454 pyrosequencing typically produces ~400 bp reads, Illumina MiSeq produces 400-700bp reads (depending on whether paired end options are used), and SOLiD produce 25–75 bp reads.<sup>[23]</sup> Historically, these read lengths were significantly shorter than the typical Sanger sequencing read length of ~750 bp, however the Illumina technology is quickly coming close to this benchmark. However, this limitation is compensated for by the much larger number of sequence reads. In 2009, pyrosequenced metagenomes generate 200–500 megabases, and Illumina platforms generate around 20–50 gigabases, but these outputs have increased by orders of magnitude in recent years.<sup>[24]</sup>



Flow diagram of a typical metagenome project<sup>[19]</sup>

An emerging approach combines shotgun sequencing and chromosome conformation capture (Hi-C), which measures the proximity of any two DNA sequences within the same cell, to guide microbial genome assembly.<sup>[25]</sup> Long read sequencing technologies, including PacBio RSII and PacBio Sequel by Pacific Biosciences, and Nanopore MinION, GridION, PromethION by Oxford Nanopore Technologies, is another choice to get long shotgun sequencing reads that should make ease in assembling process.<sup>[26]</sup>

## Bioinformatics

The data generated by metagenomics experiments are both enormous and inherently noisy, containing fragmented data representing as many as 10,000 species.<sup>[1]</sup> The sequencing of the cow rumen metagenome generated 279 gigabases, or 279 billion base pairs of nucleotide sequence data,<sup>[28]</sup> while the human gut microbiome gene catalog identified 3.3 million genes assembled from 567.7 gigabases of sequence data.<sup>[29]</sup> Collecting, curating, and extracting useful biological information from datasets of this size represent significant computational challenges for researchers.<sup>[21][30][31][32]</sup>

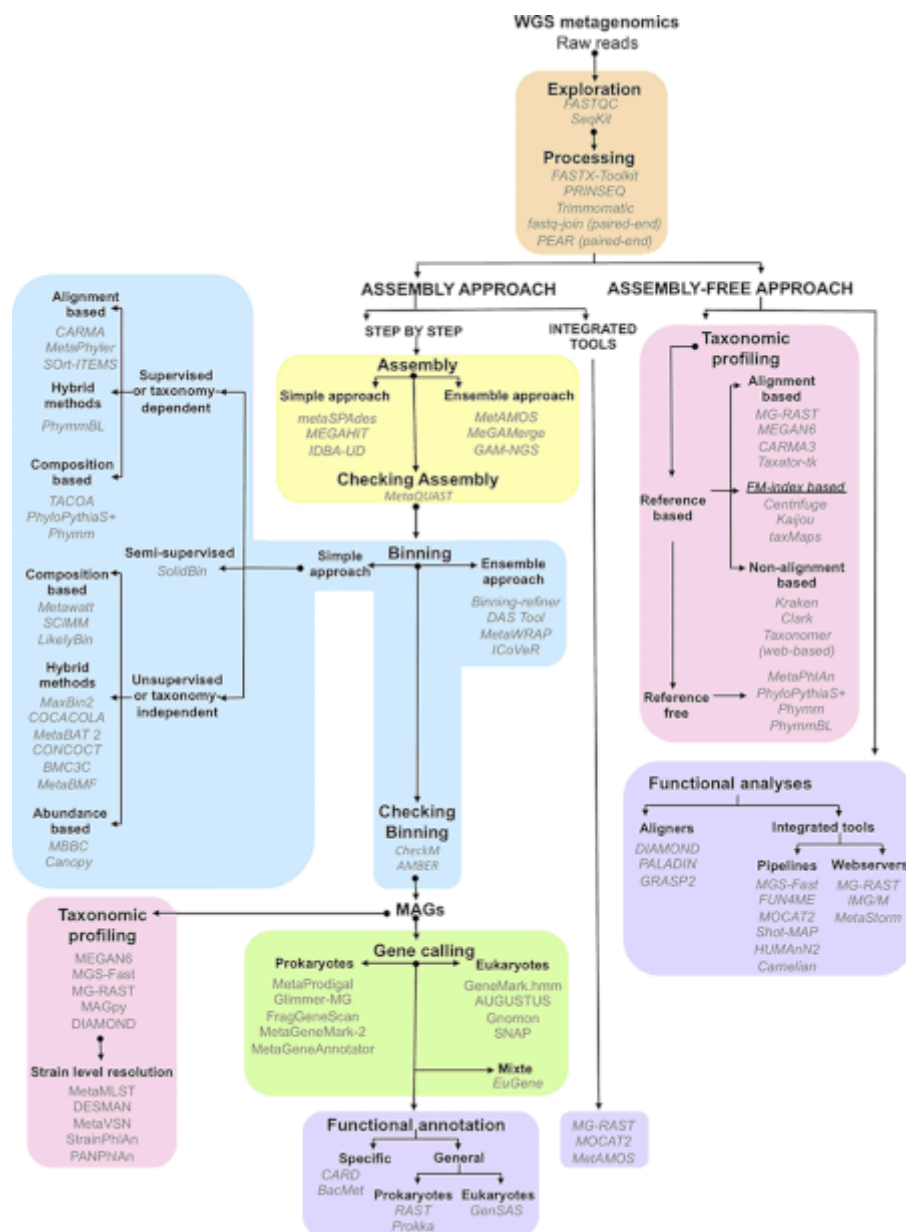
## Sequence pre-filtering

The first step of metagenomic data analysis requires the execution of certain pre-filtering steps, including the removal of redundant, low-quality sequences and sequences of probable eukaryotic origin (especially in metagenomes of human origin).<sup>[33][34]</sup> The methods available for the removal of contaminating eukaryotic

genomic DNA sequences include *Eu-Detect* and *DeConseq*.<sup>[35][36]</sup>

## Assembly

DNA sequence data from genomic and metagenomic projects are essentially the same, but genomic sequence data offers higher coverage while metagenomic data is usually highly non-redundant.<sup>[31]</sup> Furthermore, the increased use of second-generation sequencing technologies with short read lengths means that much of future metagenomic data will be error-prone. Taken in combination, these factors make the assembly of metagenomic sequence reads into genomes difficult and unreliable. Misassemblies are caused by the presence of repetitive DNA sequences that make assembly especially difficult because of the difference in the relative abundance of species present in the sample.<sup>[37]</sup> Misassemblies can also involve the combination of sequences from more than one species into chimeric contigs.<sup>[37]</sup>



Schematic representation of the main steps necessary for the analysis of whole metagenome shotgun sequencing-derived data.<sup>[27]</sup> The software related to each step is shown in *italics*.

There are several assembly programs, most of which can use information from paired-end tags in order to improve the accuracy of assemblies. Some programs, such as *Phrap* or *Celera Assembler*, were designed to be used to assemble single genomes but nevertheless produce good results when assembling metagenomic data sets.<sup>[1]</sup> Other programs, such as *Velvet assembler*, have been optimized for the shorter reads produced by second-generation sequencing through the use of de Bruijn graphs.<sup>[38][39]</sup> The use of reference genomes allows researchers to improve the assembly of the most abundant microbial species, but this approach is limited by the small subset of microbial phyla for which sequenced genomes are available.<sup>[37]</sup> After an assembly is created, an additional challenge is "metagenomic deconvolution", or determining which sequences come from which species in the sample.<sup>[40]</sup>

## Gene prediction

Metagenomic analysis pipelines use two approaches in the annotation of coding regions in the assembled contigs.<sup>[37]</sup> The first approach is to identify genes based upon homology with genes that are already publicly available in sequence databases, usually by BLAST searches. This type of approach is implemented in the program MEGAN4.<sup>[41]</sup> The second, *ab initio*, uses intrinsic features of the sequence to predict coding regions based upon gene training sets from related organisms. This is the approach taken by programs such as GeneMark<sup>[42]</sup> and GLIMMER. The main advantage of *ab initio* prediction is that it enables the detection of coding regions that lack homologs in the sequence databases; however, it is most accurate when there are large regions of contiguous genomic DNA available for comparison.<sup>[1]</sup>

## Species diversity

Gene annotations provide the "what", while measurements of species diversity provide the "who".<sup>[43]</sup> In order to connect community composition and function in metagenomes, sequences must be binned. Binning is the process of associating a particular sequence with an organism.<sup>[37]</sup> In similarity-based binning, methods such as BLAST are used to rapidly search for phylogenetic markers or otherwise similar sequences in existing public databases. This approach is implemented in MEGAN.<sup>[44]</sup> Another tool, PhymmBL, uses interpolated Markov models to assign reads.<sup>[1]</sup> MetaPhlAn (<http://huttenhower.sph.harvard.edu/metaphlan>) and AMPHORA are methods based on unique clade-specific markers for estimating organismal relative abundances with improved computational performances.<sup>[45]</sup> Other tools, like mOTUs (<https://motu-tool.org/>)<sup>[46][47]</sup> and MetaPhyler,<sup>[48]</sup> use universal marker genes to profile prokaryotic species. With the mOTUs profiler (<https://motu-tool.org/>) is possible to profile species without a reference genome, improving the estimation of microbial community diversity.<sup>[47]</sup> Recent methods, such as SLIMM (<https://github.com/seqan/slimm>), use read coverage landscape of individual reference genomes to minimize false-positive hits and get reliable relative abundances.<sup>[49]</sup> In composition based binning, methods use intrinsic features of the sequence, such as oligonucleotide frequencies or codon usage bias.<sup>[1]</sup> Once sequences are binned, it is possible to carry out comparative analysis of diversity and richness.

## Data integration

The massive amount of exponentially growing sequence data is a daunting challenge that is complicated by the complexity of the metadata associated with metagenomic projects. Metadata includes detailed information about the three-dimensional (including depth, or height) geography and environmental features of the sample, physical data about the sample site, and the methodology of the sampling.<sup>[31]</sup> This information is necessary both to ensure replicability and to enable downstream analysis. Because of its importance, metadata and collaborative data review and curation require standardized data formats located in specialized databases, such as the Genomes OnLine Database (GOLD).<sup>[50]</sup>

Several tools have been developed to integrate metadata and sequence data, allowing downstream comparative analyses of different datasets using a number of ecological indices. In 2007, Folker Meyer and Robert Edwards and a team at Argonne National Laboratory and the University of Chicago released the Metagenomics Rapid Annotation using Subsystem Technology server (MG-RAST) a community resource for metagenome data set analysis.<sup>[51]</sup> As of June 2012 over 14.8 terabases ( $14 \times 10^{12}$  bases) of DNA have been analyzed, with more than 10,000 public data sets freely available for comparison within MG-RAST. Over 8,000 users now have submitted a total of 50,000 metagenomes to MG-RAST. The Integrated Microbial Genomes/Metagenomes (IMG/M) system also provides a collection of tools for functional analysis of microbial communities based on their metagenome sequence, based upon reference isolate genomes included from the Integrated Microbial Genomes (<http://img.jgi.doe.gov/cgi-bin/w/main.cgi>) (IMG) system and the Genomic Encyclopedia of Bacteria and Archaea (GEBA) (<http://jgi.doe.gov/programs/GEBA/index.html>) project.<sup>[52]</sup>

One of the first standalone tools for analysing high-throughput metagenome shotgun data was MEGAN (MEta Genome ANalyzer).<sup>[41][44]</sup> A first version of the program was used in 2005 to analyse the metagenomic context of DNA sequences obtained from a mammoth bone.<sup>[17]</sup> Based on a BLAST comparison against a reference database, this tool performs both taxonomic and functional binning, by placing the reads onto the nodes of the NCBI taxonomy using a simple lowest common ancestor (LCA) algorithm or onto the nodes of the SEED ([http://www.theseed.org/wiki/Main\\_Page](http://www.theseed.org/wiki/Main_Page)) or KEGG classifications, respectively.<sup>[53]</sup>

With the advent of fast and inexpensive sequencing instruments, the growth of databases of DNA sequences is now exponential (e.g., the NCBI GenBank database <sup>[54]</sup>). Faster and efficient tools are needed to keep pace with the high-throughput sequencing, because the BLAST-based approaches such as MGRAST or MEGAN run slowly to annotate large samples (e.g., several hours to process a small/medium size dataset/sample <sup>[55]</sup>). Thus, ultra-fast classifiers have recently emerged, thanks to more affordable powerful servers. These tools can perform the taxonomic annotation at extremely high speed, for example CLARK <sup>[56]</sup> (according to CLARK's authors, it can classify accurately "32 million metagenomic short reads per minute"). At such a speed, a very large dataset/sample of a billion short reads can be processed in about 30 minutes.

With the increasing availability of samples containing ancient DNA and due to the uncertainty associated with the nature of those samples (ancient DNA damage),<sup>[57]</sup> a fast tool capable of producing conservative similarity estimates has been made available. According to FALCON's authors, it can use relaxed thresholds and edit distances without affecting the memory and speed performance.

## Comparative metagenomics

Comparative analyses between metagenomes can provide additional insight into the function of complex microbial communities and their role in host health.<sup>[58]</sup> Pairwise or multiple comparisons between metagenomes can be made at the level of sequence composition (comparing GC-content or genome size), taxonomic diversity, or functional complement. Comparisons of population structure and phylogenetic diversity can be made on the basis of 16S and other phylogenetic marker genes, or—in the case of low-diversity communities—by genome reconstruction from the metagenomic dataset.<sup>[59]</sup> Functional comparisons between metagenomes may be made by comparing sequences against reference databases such as COG or KEGG, and tabulating the abundance by category and evaluating any differences for statistical significance.<sup>[53]</sup> This gene-centric approach emphasizes the functional complement of the *community* as a whole rather than taxonomic groups, and shows that the functional complements are analogous under similar environmental conditions.<sup>[59]</sup> Consequently, metadata on the environmental context of the metagenomic sample is especially important in comparative analyses, as it provides researchers with the ability to study the effect of habitat upon community structure and function.<sup>[1]</sup>

Additionally, several studies have also utilized oligonucleotide usage patterns to identify the differences across diverse microbial communities. Examples of such methodologies include the dinucleotide relative abundance approach by Willner et al.<sup>[60]</sup> and the HabiSign approach of Ghosh et al.<sup>[61]</sup> This latter study also indicated that differences in tetranucleotide usage patterns can be used to identify genes (or metagenomic reads) originating from specific habitats. Additionally some methods as TriageTools<sup>[62]</sup> or Compareads<sup>[63]</sup> detect similar reads between two read sets. The similarity measure they apply on reads is based on a number of identical words of length  $k$  shared by pairs of reads.

A key goal in comparative metagenomics is to identify microbial group(s) which are responsible for conferring specific characteristics to a given environment. However, due to issues in the sequencing technologies artifacts need to be accounted for like in metagenomeSeq.<sup>[30]</sup> Others have characterized inter-microbial interactions between the resident microbial groups. A GUI-based comparative metagenomic

analysis application called Community-Analyzer has been developed by Kuntal et al. <sup>[64]</sup> which implements a correlation-based graph layout algorithm that not only facilitates a quick visualization of the differences in the analyzed microbial communities (in terms of their taxonomic composition), but also provides insights into the inherent inter-microbial interactions occurring therein. Notably, this layout algorithm also enables grouping of the metagenomes based on the probable inter-microbial interaction patterns rather than simply comparing abundance values of various taxonomic groups. In addition, the tool implements several interactive GUI-based functionalities that enable users to perform standard comparative analyses across microbiomes.

## Data analysis

---

### Community metabolism

In many bacterial communities, natural or engineered (such as bioreactors), there is significant division of labor in metabolism (Syntrophy), during which the waste products of some organisms are metabolites for others.<sup>[65]</sup> In one such system, the methanogenic bioreactor, functional stability requires the presence of several syntrophic species (Syntrophobacterales and Synergistia) working together in order to turn raw resources into fully metabolized waste (methane).<sup>[66]</sup> Using comparative gene studies and expression experiments with microarrays or proteomics researchers can piece together a metabolic network that goes beyond species boundaries. Such studies require detailed knowledge about which versions of which proteins are coded by which species and even by which strains of which species. Therefore, community genomic information is another fundamental tool (with metabolomics and proteomics) in the quest to determine how metabolites are transferred and transformed by a community.<sup>[67]</sup>

### Metatranscriptomics

Metagenomics allows researchers to access the functional and metabolic diversity of microbial communities, but it cannot show which of these processes are active.<sup>[59]</sup> The extraction and analysis of metagenomic mRNA (the **metatranscriptome**) provides information on the regulation and expression profiles of complex communities. Because of the technical difficulties (the short half-life of mRNA, for example) in the collection of environmental RNA there have been relatively few *in situ* metatranscriptomic studies of microbial communities to date.<sup>[59]</sup> While originally limited to microarray technology, metatranscriptomics studies have made use of transcriptomics technologies to measure whole-genome expression and quantification of a microbial community,<sup>[59]</sup> first employed in analysis of ammonia oxidation in soils.<sup>[68]</sup>

### Viruses

Metagenomic sequencing is particularly useful in the study of viral communities. As viruses lack a shared universal phylogenetic marker (as 16S RNA for bacteria and archaea, and 18S RNA for eukarya), the only way to access the genetic diversity of the viral community from an environmental sample is through metagenomics. Viral metagenomes (also called viromes) should thus provide more and more information about viral diversity and evolution.<sup>[69][70][71][72][73]</sup> For example, a metagenomic pipeline called Giant Virus Finder showed the first evidence of existence of giant viruses in a saline desert<sup>[74]</sup> and in Antarctic dry valleys.<sup>[75]</sup>

## Applications

---



Metagenomics has the potential to advance knowledge in a wide variety of fields. It can also be applied to solve practical challenges in medicine, engineering, agriculture, sustainability and ecology.<sup>[31][76]</sup>

## Agriculture

The soils in which plants grow are inhabited by microbial communities, with one gram of soil containing around  $10^9$ - $10^{10}$  microbial cells which comprise about one gigabase of sequence information.<sup>[77][78]</sup> The microbial communities which inhabit soils are some of the most complex known to science, and remain poorly understood despite their economic importance.<sup>[79]</sup> Microbial consortia perform a wide variety of ecosystem services necessary for plant growth, including fixing atmospheric nitrogen, nutrient cycling, disease suppression, and sequester iron and other metals.<sup>[80]</sup> Functional metagenomics strategies are being used to explore the interactions between plants and microbes through cultivation-independent study of these microbial communities.<sup>[81][82]</sup> By allowing insights into the role of previously uncultivated or rare community members in nutrient cycling and the promotion of plant growth, metagenomic approaches can contribute to improved disease detection in crops and livestock and the adaptation of enhanced farming practices which improve crop health by harnessing the relationship between microbes and plants.<sup>[31]</sup>

## Biofuel

Biofuels are fuels derived from biomass conversion, as in the conversion of cellulose contained in corn stalks, switchgrass, and other biomass into cellulosic ethanol.<sup>[31]</sup> This process is dependent upon microbial consortia (association) that transform the cellulose into sugars, followed by the fermentation of the sugars into ethanol. Microbes also produce a variety of sources of bioenergy including methane and hydrogen.<sup>[31]</sup>

The efficient industrial-scale deconstruction of biomass requires novel enzymes with higher productivity and lower cost.<sup>[28]</sup> Metagenomic approaches to the analysis of complex microbial communities allow the targeted screening of enzymes with industrial applications in biofuel production, such as glycoside hydrolases.<sup>[83]</sup> Furthermore, knowledge of how these microbial communities function is required to control them, and metagenomics is a key tool in their understanding. Metagenomic approaches allow comparative analyses between convergent microbial systems like biogas fermenters<sup>[84]</sup> or insect herbivores such as the fungus garden of the leafcutter ants.<sup>[85]</sup>

## Biotechnology

Microbial communities produce a vast array of biologically active chemicals that are used in competition and communication.<sup>[80]</sup> Many of the drugs in use today were originally uncovered in microbes; recent progress in mining the rich genetic resource of non-culturable microbes has led to the discovery of new genes, enzymes, and natural products.<sup>[59][86]</sup> The application of metagenomics has allowed the development of commodity and fine chemicals, agrochemicals and pharmaceuticals where the benefit of enzyme-catalyzed chiral synthesis is increasingly recognized.<sup>[87]</sup>

Two types of analysis are used in the bioprospecting of metagenomic data: function-driven screening for an expressed trait, and sequence-driven screening for DNA sequences of interest.<sup>[88]</sup> Function-driven analysis seeks to identify clones expressing a desired trait or useful activity, followed by biochemical characterization and sequence analysis. This approach is limited by availability of a suitable screen and the requirement that the desired trait be expressed in the host cell. Moreover, the low rate of discovery (less than one per 1,000 clones screened) and its labor-intensive nature further limit this approach.<sup>[89]</sup> In contrast, sequence-driven analysis uses conserved DNA sequences to design PCR primers to screen clones for the sequence of interest.<sup>[88]</sup> In comparison to cloning-based approaches, using a sequence-only approach

further reduces the amount of bench work required. The application of massively parallel sequencing also greatly increases the amount of sequence data generated, which require high-throughput bioinformatic analysis pipelines.<sup>[89]</sup> The sequence-driven approach to screening is limited by the breadth and accuracy of gene functions present in public sequence databases. In practice, experiments make use of a combination of both functional and sequence-based approaches based upon the function of interest, the complexity of the sample to be screened, and other factors.<sup>[89][90]</sup> An example of success using metagenomics as a biotechnology for drug discovery is illustrated with the malacidin antibiotics.<sup>[91]</sup>

## Ecology

Metagenomics can provide valuable insights into the functional ecology of environmental communities.<sup>[92]</sup> Metagenomic analysis of the bacterial consortia found in the defecations of Australian sea lions suggests that nutrient-rich sea lion faeces may be an important nutrient source for coastal ecosystems. This is because the bacteria that are expelled simultaneously with the defecations are adept at breaking down the nutrients in the faeces into a bioavailable form that can be taken up into the food chain.<sup>[93]</sup>

DNA sequencing can also be used more broadly to identify species present in a body of water,<sup>[94]</sup> debris filtered from the air, sample of dirt, or even animal's faeces.<sup>[95]</sup> This can establish the range of invasive species and endangered species, and track seasonal populations.

## Environmental remediation

Metagenomics can improve strategies for monitoring the impact of pollutants on ecosystems and for cleaning up contaminated environments. Increased understanding of how microbial communities cope with pollutants improves assessments of the potential of contaminated sites to recover from pollution and increases the chances of bioaugmentation or biostimulation trials to succeed.<sup>[96]</sup>



Metagenomics allows the study of microbial communities like those present in this stream receiving acid drainage from surface coal mining.

## Gut microbe characterization

Microbial communities play a key role in preserving human health, but their composition and the mechanism by which they do so remains mysterious.<sup>[97]</sup> Metagenomic sequencing is being used to characterize the microbial communities from 15–18 body sites from at least 250 individuals. This is part of the Human Microbiome initiative with primary goals to determine if there is a core human microbiome, to understand the changes in the human microbiome that can be correlated with human health, and to develop new technological and bioinformatics tools to support these goals.<sup>[98]</sup>

Another medical study as part of the MetaHit (Metagenomics of the Human Intestinal Tract) project consisted of 124 individuals from Denmark and Spain consisting of healthy, overweight, and irritable bowel disease patients. The study attempted to categorize the depth and phylogenetic diversity of gastrointestinal bacteria. Using Illumina GA sequence data and SOAPdenovo, a de Bruijn graph-based tool specifically designed for assembly short reads, they were able to generate 6.58 million contigs greater than 500 bp for a total contig length of 10.3 Gb and a N50 length of 2.2 kb.

The study demonstrated that two bacterial divisions, Bacteroidetes and Firmicutes, constitute over 90% of the known phylogenetic categories that dominate distal gut bacteria. Using the relative gene frequencies found within the gut these researchers identified 1,244 metagenomic clusters that are critically important for the health of the intestinal tract. There are two types of functions in these range clusters: housekeeping and those specific to the intestine. The housekeeping gene clusters are required in all bacteria and are often major players in the main metabolic pathways including central carbon metabolism and amino acid synthesis. The gut-specific functions include adhesion to host proteins and the harvesting of sugars from globoseries glycolipids. Patients with irritable bowel syndrome were shown to exhibit 25% fewer genes and lower bacterial diversity than individuals not suffering from irritable bowel syndrome indicating that changes in patients' gut biome diversity may be associated with this condition.

While these studies highlight some potentially valuable medical applications, only 31–48.8% of the reads could be aligned to 194 public human gut bacterial genomes and 7.6–21.2% to bacterial genomes available in GenBank which indicates that there is still far more research necessary to capture novel bacterial genomes.<sup>[99]</sup>

In the Human Microbiome Project (HMP), gut microbial communities were assayed using high-throughput DNA sequencing. HMP showed that, unlike individual microbial species, many metabolic processes were present among all body habitats with varying frequencies. Microbial communities of 649 metagenomes drawn from seven primary body sites on 102 individuals were studied as part of the human microbiome project. The metagenomic analysis revealed variations in niche specific abundance among 168 functional modules and 196 metabolic pathways within the microbiome. These included glycosaminoglycan degradation in the gut, as well as phosphate and amino acid transport linked to host phenotype (vaginal pH) in the posterior fornix. The HMP has brought to light the utility of metagenomics in diagnostics and evidence-based medicine. Thus metagenomics is a powerful tool to address many of the pressing issues in the field of Personalized medicine.<sup>[100]</sup>

## Infectious disease diagnosis

Differentiating between infectious and non-infectious illness, and identifying the underlying etiology of infection, can be quite challenging. For example, more than half of cases of encephalitis remain undiagnosed, despite extensive testing using state-of-the-art clinical laboratory methods. Metagenomic sequencing shows promise as a sensitive and rapid method to diagnose infection by comparing genetic material found in a patient's sample to databases of all known microscopic human pathogens and thousands of other bacterial, viral, fungal, and parasitic organisms and databases on antimicrobial resistances gene sequences with associated clinical phenotypes.

## See also

---

- Binning
- Epidemiology and sewage
- Metaproteomics
- Microbial ecology
- Pathogenomics

## References

---

1. Wooley JC, Godzik A, Friedberg I (February 2010). Bourne PE (ed.). "A primer on metagenomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829047>). *PLOS Computational Biology*. **6** (2): e1000667. Bibcode:2010PLSCB...6E0667W (<https://ui.adsab>

- s.harvard.edu/abs/2010PLSCB...6E0667W). doi:10.1371/journal.pcbi.1000667 (<https://doi.org/10.1371%2Fjournal.pcbi.1000667>). PMC 2829047 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829047>). PMID 20195499 (<https://pubmed.ncbi.nlm.nih.gov/20195499>).
2. Hugenholtz P, Goebel BM, Pace NR (September 1998). "Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107498>). *Journal of Bacteriology*. **180** (18): 4765–74. doi:10.1128/JB.180.18.4765-4774.1998 (<https://doi.org/10.1128%2FJB.180.18.4765-4774.1998>). PMC 107498 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107498>). PMID 9733676 (<https://pubmed.ncbi.nlm.nih.gov/9733676>).
  3. Marco, D, ed. (2011). *Metagenomics: Current Innovations and Future Trends*. Caister Academic Press. ISBN 978-1-904455-87-5.
  4. Eisen JA (March 2007). "Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1821061>). *PLOS Biology*. **5** (3): e82. doi:10.1371/journal.pbio.0050082 (<https://doi.org/10.1371%2Fjournal.pbio.0050082>). PMC 1821061 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1821061>). PMID 17355177 (<https://pubmed.ncbi.nlm.nih.gov/17355177>).
  5. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (October 1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products" (<https://doi.org/10.1016%2FS1074-5521%2898%2990108-9>). *Chemistry & Biology*. **5** (10): R245-9. doi:10.1016/S1074-5521(98)90108-9 (<https://doi.org/10.1016%2FS1074-5521%2898%2990108-9>). PMID 9818143 (<https://pubmed.ncbi.nlm.nih.gov/9818143>).
  6. Chen K, Pachter L (July 2005). "Bioinformatics for whole-genome shotgun sequencing of microbial communities" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1185649>). *PLOS Computational Biology*. **1** (2): 106–12. Bibcode:2005PLSCB...1...24C (<https://ui.adsabs.harvard.edu/abs/2005PLSCB...1...24C>). doi:10.1371/journal.pcbi.0010024 (<https://doi.org/10.1371%2Fjournal.pcbi.0010024>). PMC 1185649 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1185649>). PMID 16110337 (<https://pubmed.ncbi.nlm.nih.gov/16110337>).
  7. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (October 1985). "Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC391288>). *Proceedings of the National Academy of Sciences of the United States of America*. **82** (20): 6955–9. Bibcode:1985PNAS...82.6955L (<https://ui.adsabs.harvard.edu/abs/1985PNAS...82.6955L>). doi:10.1073/pnas.82.20.6955 (<https://doi.org/10.1073%2Fpnas.82.20.6955>). PMC 391288 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC391288>). PMID 2413450 (<https://pubmed.ncbi.nlm.nih.gov/2413450>).
  8. Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986). "The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences". In Marshall KC (ed.). *Advances in Microbial Ecology*. **9**. Springer US. pp. 1–55. doi:10.1007/978-1-4757-0611-6\_1 ([https://doi.org/10.1007%2F978-1-4757-0611-6\\_1](https://doi.org/10.1007%2F978-1-4757-0611-6_1)). ISBN 978-1-4757-0611-6.
  9. Schmidt TM, DeLong EF, Pace NR (July 1991). "Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC208098>). *Journal of Bacteriology*. **173** (14): 4371–8. doi:10.1128/jb.173.14.4371-4378.1991 (<https://doi.org/10.1128%2Fjb.173.14.4371-4378.1991>). PMC 208098 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC208098>). PMID 2066334 (<https://pubmed.ncbi.nlm.nih.gov/2066334>).
  10. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT (1995). "Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose". *Applied Microbiology and Biotechnology*. **43** (4): 667–74. doi:10.1007/BF00164771 (<https://doi.org/10.1007%2FBF00164771>). PMID 7546604 (<https://pubmed.ncbi.nlm.nih.gov/7546604>). S2CID 31384119 (<https://api.semanticscholar.org/CorpusID:31384119>).

11. Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (February 1996). "Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC177699>). *Journal of Bacteriology*. **178** (3): 591–9. doi:10.1128/jb.178.3.591-599.1996 (<https://doi.org/10.1128%2Fjb.178.3.591-599.1996>). PMC 177699 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC177699>). PMID 8550487 (<https://pubmed.ncbi.nlm.nih.gov/8550487>).
12. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. (October 2002). "Genomic analysis of uncultured marine viral communities" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC137870>). *Proceedings of the National Academy of Sciences of the United States of America*. **99** (22): 14250–5. Bibcode:2002PNAS...9914250B (<https://ui.adsabs.harvard.edu/abs/2002PNAS...9914250B>). doi:10.1073/pnas.202488399 (<https://doi.org/10.1073%2Fpnas.202488399>). PMC 137870 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC137870>). PMID 12384570 (<https://pubmed.ncbi.nlm.nih.gov/12384570>).
13. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. (March 2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment". *Nature*. **428** (6978): 37–43. Bibcode:2004Natur.428...37T (<https://ui.adsabs.harvard.edu/abs/2004Natur.428...37T>). doi:10.1038/nature02340 (<https://doi.org/10.1038%2Fnature02340>). PMID 14961025 (<https://pubmed.ncbi.nlm.nih.gov/14961025>). S2CID 4420754 (<https://api.semanticscholar.org/CorpusID:4420754>). (subscription required)
14. Hugenholtz P (2002). "Exploring prokaryotic diversity in the genomic era" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139013>). *Genome Biology*. **3** (2): REVIEWS0003. doi:10.1186/gb-2002-3-2-reviews0003 (<https://doi.org/10.1186%2Fgb-2002-3-2-reviews0003>). PMC 139013 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139013>). PMID 11864374 (<https://pubmed.ncbi.nlm.nih.gov/11864374>).
15. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. (April 2004). "Environmental genome shotgun sequencing of the Sargasso Sea". *Science*. **304** (5667): 66–74. Bibcode:2004Sci...304...66V (<https://ui.adsabs.harvard.edu/abs/2004Sci...304...66V>). CiteSeerX 10.1.1.124.1840 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.124.1840>). doi:10.1126/science.1093857 (<https://doi.org/10.1126%2Fscience.1093857>). PMID 15001713 (<https://pubmed.ncbi.nlm.nih.gov/15001713>). S2CID 1454587 (<https://api.semanticscholar.org/CorpusID:1454587>).
16. Yoosseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M, et al. (November 2010). "Genomic and functional adaptation in surface ocean planktonic prokaryotes" (<https://doi.org/10.1038%2Fnature09530>). *Nature*. **468** (7320): 60–6. Bibcode:2010Natur.468...60Y (<https://ui.adsabs.harvard.edu/abs/2010Natur.468...60Y>). doi:10.1038/nature09530 (<https://doi.org/10.1038%2Fnature09530>). PMID 21048761 (<https://pubmed.ncbi.nlm.nih.gov/21048761>). (subscription required)
17. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, et al. (January 2006). "Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA" (<https://semanticscholar.org/paper/fcb647962c071b7c12e2f1513daadacce467fc9b>). *Science*. **311** (5759): 392–4. Bibcode:2006Sci...311..392P (<https://ui.adsabs.harvard.edu/abs/2006Sci...311..392P>). doi:10.1126/science.1123360 (<https://doi.org/10.1126%2Fscience.1123360>). PMID 16368896 (<https://pubmed.ncbi.nlm.nih.gov/16368896>). S2CID 11238470 (<https://api.semanticscholar.org/CorpusID:11238470>).
18. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, et al. (March 2006). "Using pyrosequencing to shed light on deep mine microbial ecology" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1483832>). *BMC Genomics*. **7**: 57. doi:10.1186/1471-2164-7-57 (<https://doi.org/10.1186%2F1471-2164-7-57>). PMC 1483832 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1483832>). PMID 16549033 (<https://pubmed.ncbi.nlm.nih.gov/16549033>).

19. Thomas T, Gilbert J, Meyer F (February 2012). "Metagenomics - a guide from sampling to data analysis" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351745>). *Microbial Informatics and Experimentation*. **2** (1): 3. doi:10.1186/2042-5783-2-3 (<https://doi.org/10.1186/2042-5783-2-3>). PMC 3351745 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351745>). PMID 22587947 (<https://pubmed.ncbi.nlm.nih.gov/22587947>).
20. Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, et al. (October 2000). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage" (<https://semanticscholar.org/paper/1f388e1b54a075e03cf7096d21c583db32d66293>). *Environmental Microbiology*. **2** (5): 516–29. doi:10.1046/j.1462-2920.2000.00133.x (<https://doi.org/10.1046/j.1462-2920.2000.00133.x>). PMID 11233160 (<https://pubmed.ncbi.nlm.nih.gov/11233160>). S2CID 8267748 (<https://api.semanticscholar.org/CorpusID:8267748>).
21. Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C (May 2013). "Computational meta'omics for microbial community studies" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4039370>). *Molecular Systems Biology*. **9** (666): 666. doi:10.1038/msb.2013.22 (<https://doi.org/10.1038/msb.2013.22>). PMC 4039370 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4039370>). PMID 23670539 (<https://pubmed.ncbi.nlm.nih.gov/23670539>).
22. Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ, Chisholm SW (July 2010). Gilbert JA (ed.). "Unlocking short read sequencing for metagenomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2911387>). *PLOS ONE*. **5** (7): e11840. Bibcode:2010PLoSO...511840R (<https://ui.adsabs.harvard.edu/abs/2010PLoSO...511840R>). doi:10.1371/journal.pone.0011840 (<https://doi.org/10.1371/journal.pone.0011840>). PMC 2911387 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2911387>). PMID 20676378 (<https://pubmed.ncbi.nlm.nih.gov/20676378>).
23. Schuster SC (January 2008). "Next-generation sequencing transforms today's biology". *Nature Methods*. **5** (1): 16–8. doi:10.1038/nmeth1156 (<https://doi.org/10.1038/nmeth1156>). PMID 18165802 (<https://pubmed.ncbi.nlm.nih.gov/18165802>). S2CID 1465786 (<https://api.semanticscholar.org/CorpusID:1465786>).
24. "Metagenomics versus Moore's law" (<https://doi.org/10.1038/nmeth0909-623>). *Nature Methods*. **6** (9): 623. 2009. doi:10.1038/nmeth0909-623 (<https://doi.org/10.1038/nmeth0909-623>).
25. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. (February 2018). "Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5830445>). *Nature Communications*. **9** (1): 870. Bibcode:2018NatCo...9..870S (<https://ui.adsabs.harvard.edu/abs/2018NatCo...9..870S>). doi:10.1038/s41467-018-03317-6 (<https://doi.org/10.1038/s41467-018-03317-6>). PMC 5830445 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5830445>). PMID 29491419 (<https://pubmed.ncbi.nlm.nih.gov/29491419>).
26. Hiraoka S, Yang CC, Iwasaki W (September 2016). "Metagenomics and Bioinformatics in Microbial Ecology: Current Status and Beyond" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017796>). *Microbes and Environments*. **31** (3): 204–12. doi:10.1264/jsme2.ME16024 (<https://doi.org/10.1264/jsme2.ME16024>). PMC 5017796 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017796>). PMID 27383682 (<https://pubmed.ncbi.nlm.nih.gov/27383682>).
27. Pérez-Cobas AE, Gomez-Valero L, Buchrieser C (2020). "Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7641418>). *Microbial Genomics*. **6** (8). doi:10.1099/mgen.0.000409 (<https://doi.org/10.1099/mgen.0.000409>). PMC 7641418 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7641418>). PMID 32706331 (<https://pubmed.ncbi.nlm.nih.gov/32706331>).

28. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. (January 2011). "Metagenomic discovery of biomass-degrading genes and genomes from cow rumen" (<http://semanticscholar.org/paper/53bfb21fda6801779909fba6d4d4778dbce5119d>). *Science*. **331** (6016): 463–7. Bibcode:2011Sci...331..463H (<https://ui.adsabs.harvard.edu/abs/2011Sci...331..463H>). doi:10.1126/science.1200387 (<https://doi.org/10.1126%2Fscience.1200387>). PMID 21273488 (<https://pubmed.ncbi.nlm.nih.gov/21273488>). S2CID 36572885 (<https://api.semanticscholar.org/CorpusID:36572885>).
29. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. (March 2010). "A human gut microbial gene catalogue established by metagenomic sequencing" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3779803>). *Nature*. **464** (7285): 59–65. Bibcode:2010Natur.464...59. (<https://ui.adsabs.harvard.edu/abs/2010Natur.464...59>). doi:10.1038/nature08821 (<https://doi.org/10.1038%2Fnature08821>). PMC 3779803 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3779803>). PMID 20203603 (<https://pubmed.ncbi.nlm.nih.gov/20203603>). (subscription required)
30. Paulson JN, Stine OC, Bravo HC, Pop M (December 2013). "Differential abundance analysis for microbial marker-gene surveys" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010126>). *Nature Methods*. **10** (12): 1200–2. doi:10.1038/nmeth.2658 (<https://doi.org/10.1038%2Fnmeth.2658>). PMC 4010126 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010126>). PMID 24076764 (<https://pubmed.ncbi.nlm.nih.gov/24076764>).
31. Committee on Metagenomics: Challenges and Functional Applications, National Research Council (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* (<http://www.nap.edu/catalog/11902.html>). Washington, D.C.: The National Academies Press. doi:10.17226/11902 (<https://doi.org/10.17226%2F11902>). ISBN 978-0-309-10676-4. PMID 21678629 (<https://pubmed.ncbi.nlm.nih.gov/21678629>).
32. Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, et al. (2015). "Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4426941>). *Bioinformatics and Biology Insights*. **9**: 75–88. doi:10.4137/BBI.S12462 (<https://doi.org/10.4137%2FBBI.S12462>). PMC 4426941 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4426941>). PMID 25983555 (<https://pubmed.ncbi.nlm.nih.gov/25983555>).
33. Mende DR, Waller AS, Sunagawa S, Järvelin AI, Chan MM, Arumugam M, et al. (23 February 2012). "Assessment of metagenomic assembly using simulated next generation sequencing data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285633>). *PLOS ONE*. **7** (2): e31386. Bibcode:2012PLoSO...731386M (<https://ui.adsabs.harvard.edu/abs/2012PLoSO...731386M>). doi:10.1371/journal.pone.0031386 (<https://doi.org/10.1371%2Fjournal.pone.0031386>). PMC 3285633 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285633>). PMID 22384016 (<https://pubmed.ncbi.nlm.nih.gov/22384016>).
34. Balzer S, Malde K, Grohme MA, Jonassen I (April 2013). "Filtering duplicate reads from 454 pyrosequencing data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3605598>). *Bioinformatics*. **29** (7): 830–6. doi:10.1093/bioinformatics/btt047 (<https://doi.org/10.1093%2Fbioinformatics%2Fbtt047>). PMC 3605598 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3605598>). PMID 23376350 (<https://pubmed.ncbi.nlm.nih.gov/23376350>).
35. Mohammed MH, Chadaram S, Komanduri D, Ghosh TS, Mande SS (September 2011). "Eu-Detect: an algorithm for detecting eukaryotic sequences in metagenomic data sets". *Journal of Biosciences*. **36** (4): 709–17. doi:10.1007/s12038-011-9105-2 (<https://doi.org/10.1007%2Fs12038-011-9105-2>). PMID 21857117 (<https://pubmed.ncbi.nlm.nih.gov/21857117>). S2CID 25857874 (<https://api.semanticscholar.org/CorpusID:25857874>).



36. Schmieder R, Edwards R (March 2011). "Fast identification and removal of sequence contamination from genomic and metagenomic datasets" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3052304>). *PLOS ONE*. **6** (3): e17288. Bibcode:2011PLoSO...617288S (<https://ui.adsabs.harvard.edu/abs/2011PLoSO...617288S>). doi:10.1371/journal.pone.0017288 (<https://doi.org/10.1371%2Fjournal.pone.0017288>). PMC 3052304 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3052304>). PMID 21408061 (<https://pubmed.ncbi.nlm.nih.gov/21408061>).
37. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (December 2008). "A bioinformatician's guide to metagenomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2593568>). *Microbiology and Molecular Biology Reviews*. **72** (4): 557–78, Table of Contents. doi:10.1128/MMBR.00009-08 (<https://doi.org/10.1128%2FMMBR.00009-08>). PMC 2593568 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2593568>). PMID 19052320 (<https://pubmed.ncbi.nlm.nih.gov/19052320>).
38. Namiki T, Hachiya T, Tanaka H, Sakakibara Y (November 2012). "MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3488206>). *Nucleic Acids Research*. **40** (20): e155. doi:10.1093/nar/gks678 (<https://doi.org/10.1093%2Fnar%2Fgks678>). PMC 3488206 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3488206>). PMID 22821567 (<https://pubmed.ncbi.nlm.nih.gov/22821567>).
39. Zerbino DR, Birney E (May 2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2336801>). *Genome Research*. **18** (5): 821–9. doi:10.1101/gr.074492.107 (<https://doi.org/10.1101%2Fgr.074492.107>). PMC 2336801 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2336801>). PMID 18349386 (<https://pubmed.ncbi.nlm.nih.gov/18349386>).
40. Burton JN, Liachko I, Dunham MJ, Shendure J (May 2014). "Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4455782>). *G3*. **4** (7): 1339–46. doi:10.1534/g3.114.011825 (<https://doi.org/10.1534%2Fg3.114.011825>). PMC 4455782 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4455782>). PMID 24855317 (<https://pubmed.ncbi.nlm.nih.gov/24855317>).
41. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (September 2011). "Integrative analysis of environmental sequences using MEGAN4" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3166839>). *Genome Research*. **21** (9): 1552–60. doi:10.1101/gr.120618.111 (<https://doi.org/10.1101%2Fgr.120618.111>). PMC 3166839 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3166839>). PMID 21690186 (<https://pubmed.ncbi.nlm.nih.gov/21690186>).
42. Zhu W, Lomsadze A, Borodovsky M (July 2010). "Ab initio gene identification in metagenomic sequences" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896542>). *Nucleic Acids Research*. **38** (12): e132. doi:10.1093/nar/gkq275 (<https://doi.org/10.1093%2Fnar%2Fgkq275>). PMC 2896542 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896542>). PMID 20403810 (<https://pubmed.ncbi.nlm.nih.gov/20403810>).
43. Konopka A (November 2009). "What is microbial community ecology?" (<https://doi.org/10.1038%2Fismej.2009.88>). *The ISME Journal*. **3** (11): 1223–30. doi:10.1038/ismej.2009.88 (<https://doi.org/10.1038%2Fismej.2009.88>). PMID 19657372 (<https://pubmed.ncbi.nlm.nih.gov/19657372>).
44. Huson DH, Auch AF, Qi J, Schuster SC (March 2007). "MEGAN analysis of metagenomic data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1800929>). *Genome Research*. **17** (3): 377–86. doi:10.1101/gr.5969107 (<https://doi.org/10.1101%2Fgr.5969107>). PMC 1800929 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1800929>). PMID 17255551 (<https://pubmed.ncbi.nlm.nih.gov/17255551>).
45. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (June 2012). "Metagenomic microbial community profiling using unique clade-specific marker genes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443552>). *Nature Methods*. **9** (8): 811–4. doi:10.1038/nmeth.2066 (<https://doi.org/10.1038%2Fnmeth.2066>). PMC 3443552 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443552>). PMID 22688413 (<https://pubmed.ncbi.nlm.nih.gov/22688413>).



46. Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, et al. (December 2013). "Metagenomic species profiling using universal phylogenetic marker genes". *Nature Methods*. **10** (12): 1196–9. doi:10.1038/nmeth.2693 (<https://doi.org/10.1038/nmeth.2693>). PMID 24141494 (<https://pubmed.ncbi.nlm.nih.gov/24141494>). S2CID 7728395 (<https://api.semanticscholar.org/CorpusID:7728395>).
47. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. (March 2019). "Microbial abundance, activity and population genomic profiling with mOTUs2" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6399450>). *Nature Communications*. **10** (1): 1014. Bibcode:2019NatCo..10.1014M (<https://ui.adsabs.harvard.edu/abs/2019NatCo..10.1014M>). doi:10.1038/s41467-019-08844-4 (<https://doi.org/10.1038/s41467-019-08844-4>). PMC 6399450 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6399450>). PMID 30833550 (<https://pubmed.ncbi.nlm.nih.gov/30833550>).
48. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M (2011). "Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3194235>). *BMC Genomics*. 12 Suppl 2: S4. doi:10.1186/1471-2164-12-S2-S4 (<https://doi.org/10.1186/1471-2164-12-S2-S4>). PMC 3194235 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3194235>). PMID 21989143 (<https://pubmed.ncbi.nlm.nih.gov/21989143>).
49. Dadi TH, Renard BY, Wieler LH, Semmler T, Reinert K (2017). "SLIMM: species level identification of microorganisms from metagenomes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5372838>). *PeerJ*. **5**: e3138. doi:10.7717/peerj.3138 (<https://doi.org/10.7717/peerj.3138>). PMC 5372838 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5372838>). PMID 28367376 (<https://pubmed.ncbi.nlm.nih.gov/28367376>).
50. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. (January 2012). "The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245063>). *Nucleic Acids Research*. **40** (Database issue): D571-9. doi:10.1093/nar/gkr1100 (<https://doi.org/10.1093/nar/gkr1100>). PMC 3245063 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245063>). PMID 22135293 (<https://pubmed.ncbi.nlm.nih.gov/22135293>).
51. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. (September 2008). "The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2563014>). *BMC Bioinformatics*. **9**: 386. doi:10.1186/1471-2105-9-386 (<https://doi.org/10.1186/1471-2105-9-386>). PMC 2563014 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2563014>). PMID 18803844 (<https://pubmed.ncbi.nlm.nih.gov/18803844>).
52. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Grechkin Y, et al. (January 2012). "IMG/M: the integrated metagenome data management and comparative analysis system" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245048>). *Nucleic Acids Research*. **40** (Database issue): D123-9. doi:10.1093/nar/gkr975 (<https://doi.org/10.1093/nar/gkr975>). PMC 3245048 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245048>). PMID 22086953 (<https://pubmed.ncbi.nlm.nih.gov/22086953>).
53. Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, et al. (February 2011). "Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3044276>). *BMC Bioinformatics*. 12 Suppl 1: S21. doi:10.1186/1471-2105-12-S1-S21 (<https://doi.org/10.1186/1471-2105-12-S1-S21>). PMC 3044276 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3044276>). PMID 21342551 (<https://pubmed.ncbi.nlm.nih.gov/21342551>).
54. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (January 2013). "GenBank" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531190>). *Nucleic Acids Research*. **41** (Database issue): D36-42. doi:10.1093/nar/gks1195 (<https://doi.org/10.1093/nar/gks1195>). PMC 3531190 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531190>). PMID 23193287 (<https://pubmed.ncbi.nlm.nih.gov/23193287>).

55. Bazinet AL, Cummings MP (May 2012). "A comparative evaluation of sequence classification programs" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3428669>). *BMC Bioinformatics*. **13**: 92. doi:10.1186/1471-2105-13-92 (<https://doi.org/10.1186%2F1471-2105-13-92>). PMC 3428669 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3428669>). PMID 22574964 (<https://pubmed.ncbi.nlm.nih.gov/22574964>).
56. Ounit R, Wanamaker S, Close TJ, Lonardi S (March 2015). "CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4428112>). *BMC Genomics*. **16**: 236. doi:10.1186/s12864-015-1419-2 (<https://doi.org/10.1186%2Fs12864-015-1419-2>). PMC 4428112 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4428112>). PMID 25879410 (<https://pubmed.ncbi.nlm.nih.gov/25879410>).
57. Pratas D, Pinho AJ, Silva RM, Rodrigues JM, Hosseini M, Caetano T, Ferreira PJ (February 2018). "FALCON: a method to infer metagenomic composition of ancient DNA". *bioRxiv* 10.1101/267179 (<https://doi.org/10.1101%2F267179>).
58. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, et al. (August 2007). "Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533590>). *DNA Research*. **14** (4): 169–81. doi:10.1093/dnares/dsm018 (<https://doi.org/10.1093%2Fdnares%2Fdsm018>). PMC 2533590 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533590>). PMID 17916580 (<https://pubmed.ncbi.nlm.nih.gov/17916580>).
59. Simon C, Daniel R (February 2011). "Metagenomic analyses: past and future trends" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3067235>). *Applied and Environmental Microbiology*. **77** (4): 1153–61. doi:10.1128/AEM.02345-10 (<https://doi.org/10.1128%2FAEM.02345-10>). PMC 3067235 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3067235>). PMID 21169428 (<https://pubmed.ncbi.nlm.nih.gov/21169428>).
60. Willner D, Thurber RV, Rohwer F (July 2009). "Metagenomic signatures of 86 microbial and viral metagenomes". *Environmental Microbiology*. **11** (7): 1752–66. doi:10.1111/j.1462-2920.2009.01901.x (<https://doi.org/10.1111%2Fj.1462-2920.2009.01901.x>). PMID 19302541 (<https://pubmed.ncbi.nlm.nih.gov/19302541>).
61. Ghosh TS, Mohammed MH, Rajasingh H, Chadaram S, Mande SS (2011). "HabiSign: a novel approach for comparison of metagenomes and rapid identification of habitat-specific sequences" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3278849>). *BMC Bioinformatics*. **12** Suppl 13 (Supplement 13): S9. doi:10.1186/1471-2105-12-s13-s9 (<https://doi.org/10.1186%2F1471-2105-12-s13-s9>). PMC 3278849 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3278849>). PMID 22373355 (<https://pubmed.ncbi.nlm.nih.gov/22373355>).
62. Fimereli D, Detours V, Konopka T (April 2013). "TriageTools: tools for partitioning and prioritizing analysis of high-throughput sequencing data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627586>). *Nucleic Acids Research*. **41** (7): e86. doi:10.1093/nar/gkt094 (<https://doi.org/10.1093%2Fnar%2Fgkt094>). PMC 3627586 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627586>). PMID 23408855 (<https://pubmed.ncbi.nlm.nih.gov/23408855>).
63. Maillet N, Lemaitre C, Chikhi R, Lavenier D, Peterlongo P (2012). "Compareads: comparing huge metagenomic experiments" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526429>). *BMC Bioinformatics*. **13** Suppl 19 (Suppl 19): S10. doi:10.1186/1471-2105-13-S19-S10 (<https://doi.org/10.1186%2F1471-2105-13-S19-S10>). PMC 3526429 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526429>). PMID 23282463 (<https://pubmed.ncbi.nlm.nih.gov/23282463>).
64. Kuntal BK, Ghosh TS, Mande SS (October 2013). "Community-analyzer: a platform for visualizing and comparing microbial community structure across microbiomes" (<https://doi.org/10.1016%2Fj.ygeno.2013.08.004>). *Genomics*. **102** (4): 409–18. doi:10.1016/j.ygeno.2013.08.004 (<https://doi.org/10.1016%2Fj.ygeno.2013.08.004>). PMID 23978768 (<https://pubmed.ncbi.nlm.nih.gov/23978768>).

65. Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K, et al. (March 2011). "Bacterial community structures are unique and resilient in full-scale bioenergy systems" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3053989>). *Proceedings of the National Academy of Sciences of the United States of America*. **108** (10): 4158–63. Bibcode:2011PNAS..108.4158W (<https://ui.adsabs.harvard.edu/abs/2011PNAS..108.4158W>). doi:10.1073/pnas.1015676108 (<https://doi.org/10.1073%2Fpnas.1015676108>). PMC 3053989 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3053989>). PMID 21368115 (<https://pubmed.ncbi.nlm.nih.gov/21368115>).
66. McInerney MJ, Sieber JR, Gunsalus RP (December 2009). "Syntrophy in anaerobic global carbon cycles" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2790021>). *Current Opinion in Biotechnology*. **20** (6): 623–32. doi:10.1016/j.copbio.2009.10.001 (<https://doi.org/10.1016%2Fj.copbio.2009.10.001>). PMC 2790021 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2790021>). PMID 19897353 (<https://pubmed.ncbi.nlm.nih.gov/19897353>).
67. Klitgord N, Segrè D (August 2011). "Ecosystems biology of microbial metabolism". *Current Opinion in Biotechnology*. **22** (4): 541–6. doi:10.1016/j.copbio.2011.04.018 (<https://doi.org/10.1016%2Fj.copbio.2011.04.018>). PMID 21592777 (<https://pubmed.ncbi.nlm.nih.gov/21592777>).
68. Leininger S, Urich T, Schlöter M, Schwark L, Qi J, Nicol GW, et al. (August 2006). "Archaea predominate among ammonia-oxidizing prokaryotes in soils". *Nature*. **442** (7104): 806–9. Bibcode:2006Natur.442..806L (<https://ui.adsabs.harvard.edu/abs/2006Natur.442..806L>). doi:10.1038/nature04983 (<https://doi.org/10.1038%2Fnature04983>). PMID 16915287 (<https://pubmed.ncbi.nlm.nih.gov/16915287>). S2CID 4380804 (<https://api.semanticscholar.org/CorpusID:4380804>).
69. Paez-Espino D, Elie-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. (August 2016). "Uncovering Earth's virome" (<http://www.escholarship.org/uc/item/4zh090xt>). *Nature*. **536** (7617): 425–30. Bibcode:2016Natur.536..425P (<https://ui.adsabs.harvard.edu/abs/2016Natur.536..425P>). doi:10.1038/nature19094 (<https://doi.org/10.1038%2Fnature19094>). PMID 27533034 (<https://pubmed.ncbi.nlm.nih.gov/27533034>). S2CID 4466854 (<https://api.semanticscholar.org/CorpusID:4466854>).
70. Paez-Espino D, Chen IA, Palaniappan K, Ratner A, Chu K, Szeto E, et al. (January 2017). "IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210529>). *Nucleic Acids Research*. **45** (D1): D457–D465. doi:10.1093/nar/gkw1030 (<https://doi.org/10.1093%2Fnar%2Fgkw1030>). PMC 5210529 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210529>). PMID 27799466 (<https://pubmed.ncbi.nlm.nih.gov/27799466>).
71. Paez-Espino D, Roux S, Chen IA, Palaniappan K, Ratner A, Chu K, et al. (January 2019). "IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6323928>). *Nucleic Acids Research*. **47** (D1): D678–D686. doi:10.1093/nar/gky1127 (<https://doi.org/10.1093%2Fnar%2Fgky1127>). PMC 6323928 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6323928>). PMID 30407573 (<https://pubmed.ncbi.nlm.nih.gov/30407573>).
72. Paez-Espino D, Pavlopoulos GA, Ivanova NN, Kyrpides NC (August 2017). "Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data" (<https://escholarship.org/content/qt1549t4d3/qt1549t4d3.pdf?t=p2n6i2>) (PDF). *Nature Protocols*. **12** (8): 1673–1682. doi:10.1038/nprot.2017.063 (<https://doi.org/10.1038%2Fnprot.2017.063>). PMID 28749930 (<https://pubmed.ncbi.nlm.nih.gov/28749930>). S2CID 2127494 (<https://api.semanticscholar.org/CorpusID:2127494>).
73. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV (January 2010). "New dimensions of the virus world discovered through metagenomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293453>). *Trends in Microbiology*. **18** (1): 11–9. doi:10.1016/j.tim.2009.11.003 (<https://doi.org/10.1016%2Fj.tim.2009.11.003>). PMC 3293453 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293453>). PMID 19942437 (<https://pubmed.ncbi.nlm.nih.gov/19942437>).

74. Kerepesi C, Grolmusz V (March 2016). "Giant viruses of the Kutch Desert". *Archives of Virology*. **161** (3): 721–4. arXiv:1410.1278 (<https://arxiv.org/abs/1410.1278>). doi:10.1007/s00705-015-2720-8 (<https://doi.org/10.1007%2Fs00705-015-2720-8>). PMID 26666442 (<https://pubmed.ncbi.nlm.nih.gov/26666442>). S2CID 13145926 (<https://api.semanticscholar.org/CorpusID:13145926>).
75. Kerepesi C, Grolmusz V (June 2017). "The "Giant Virus Finder" discovers an abundance of giant viruses in the Antarctic dry valleys". *Archives of Virology*. **162** (6): 1671–1676. arXiv:1503.05575 (<https://arxiv.org/abs/1503.05575>). doi:10.1007/s00705-017-3286-4 (<https://doi.org/10.1007%2Fs00705-017-3286-4>). PMID 28247094 (<https://pubmed.ncbi.nlm.nih.gov/28247094>). S2CID 1925728 (<https://api.semanticscholar.org/CorpusID:1925728>).
76. Copeland CS (September–October 2017). "The World Within Us" ([http://claudiacopeland.com/uploads/3/5/5/6/35560346/hjno\\_the\\_world\\_within\\_us.pdf](http://claudiacopeland.com/uploads/3/5/5/6/35560346/hjno_the_world_within_us.pdf)) (PDF). *Healthcare Journal of New Orleans*: 21–26.
77. Jansson J (2011). "Towards "Tera-Terra": Terabase Sequencing of Terrestrial Metagenomes Print E-mail" (<https://web.archive.org/web/20120331040553/http://www.microbemagazine.org/index.php/07-2011-home/3553-towards-tera-terra-terabase-sequencing-of-terrestrial-metagenomes>). *Microbe*. **6** (7). p. 309. Archived from the original (<http://www.microbemagazine.org/index.php/07-2011-home/3553-towards-tera-terra-terabase-sequencing-of-terrestrial-metagenomes>) on 31 March 2012.
78. Vogel TM, Simonet P, Jansson JK, Hirsch PR, Tiedje JM, Van Elsas JD, Bailey MJ, Nalin R, Philippot L (2009). "TerraGenome: A consortium for the sequencing of a soil metagenome" (<https://doi.org/10.1038%2Fnrmicro2119>). *Nature Reviews Microbiology*. **7** (4): 252. doi:10.1038/nrmicro2119 (<https://doi.org/10.1038%2Fnrmicro2119>).
79. "TerraGenome Homepage" (<http://www.terragenome.org/>). *TerraGenome international sequencing consortium*. Retrieved 30 December 2011.
80. Committee on Metagenomics: Challenges and Functional Applications, National Research Council (2007). *Understanding Our Microbial Planet: The New Science of Metagenomics* ([http://dels-old.nas.edu/dels/rpt\\_briefs/metagenomics\\_final.pdf](http://dels-old.nas.edu/dels/rpt_briefs/metagenomics_final.pdf)) (PDF). The National Academies Press.
81. Charles T (2010). "The Potential for Investigation of Plant-microbe Interactions Using Metagenomics Methods". *Metagenomics: Theory, Methods and Applications*. Caister Academic Press. ISBN 978-1-904455-54-7.
82. Bringel F, Couée I (22 May 2015). "Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4440916>). *Frontiers in Microbiology*. **6**: 486. doi:10.3389/fmicb.2015.00486 (<https://doi.org/10.3389%2Ffmicb.2015.00486>). PMC 4440916 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4440916>). PMID 26052316 (<https://pubmed.ncbi.nlm.nih.gov/26052316>).
83. Li LL, McCorkle SR, Monchy S, Taghavi S, van der Lelie D (May 2009). "Bioprospecting metagenomes: glycosyl hydrolases for converting biomass" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2694162>). *Biotechnology for Biofuels*. **2**: 10. doi:10.1186/1754-6834-2-10 (<https://doi.org/10.1186%2F1754-6834-2-10>). PMC 2694162 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2694162>). PMID 19450243 (<https://pubmed.ncbi.nlm.nih.gov/19450243>).
84. Jaenicke S, Ander C, Bekel T, Bisdorf R, Dröge M, Gartemann KH, et al. (January 2011). Aziz RK (ed.). "Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3027613>). *PLOS ONE*. **6** (1): e14519. Bibcode:2011PLoSO...614519J (<https://ui.adsabs.harvard.edu/abs/2011PLoSO...614519J>). doi:10.1371/journal.pone.0014519 (<https://doi.org/10.1371%2Fjournal.pone.0014519>). PMC 3027613 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3027613>). PMID 21297863 (<https://pubmed.ncbi.nlm.nih.gov/21297863>).

85. Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA, et al. (September 2010). Sonnenburg J (ed.). "An insect herbivore microbiome with high plant biomass-degrading capacity" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2944797>). *PLOS Genetics*. **6** (9): e1001129. doi:10.1371/journal.pgen.1001129 (<https://doi.org/10.1371%2Fjournal.pgen.1001129>). PMC 2944797 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2944797>). PMID 20885794 (<https://pubmed.ncbi.nlm.nih.gov/20885794>).
86. Simon C, Daniel R (November 2009). "Achievements and new knowledge unraveled by metagenomic approaches" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2773367>). *Applied Microbiology and Biotechnology*. **85** (2): 265–76. doi:10.1007/s00253-009-2233-z (<https://doi.org/10.1007%2Fs00253-009-2233-z>). PMC 2773367 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2773367>). PMID 19760178 (<https://pubmed.ncbi.nlm.nih.gov/19760178>).
87. Wong D (2010). "Applications of Metagenomics for Industrial Bioproducts". *Metagenomics: Theory, Methods and Applications*. Caister Academic Press. ISBN 978-1-904455-54-7.
88. Schloss PD, Handelsman J (June 2003). "Biotechnological prospects from metagenomics" (<https://web.archive.org/web/20160304024004/http://handelsmanlab.sites.yale.edu/sites/default/files/Biotechnological%20prospects%20from%20metagenomics.pdf>) (PDF). *Current Opinion in Biotechnology*. **14** (3): 303–10. doi:10.1016/S0958-1669(03)00067-3 (<https://doi.org/10.1016%2FS0958-1669%2803%2900067-3>). PMID 12849784 (<https://pubmed.ncbi.nlm.nih.gov/12849784>). Archived from the original (<http://handelsmanlab.sites.yale.edu/sites/default/files/Biotechnological%20prospects%20from%20metagenomics.pdf>) (PDF) on 4 March 2016. Retrieved 20 January 2012.
89. Kakirde KS, Parsley LC, Liles MR (November 2010). "Size Does Matter: Application-driven Approaches for Soil Metagenomics" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976544>). *Soil Biology & Biochemistry*. **42** (11): 1911–1923. doi:10.1016/j.soilbio.2010.07.021 (<https://doi.org/10.1016%2Fj.soilbio.2010.07.021>). PMC 2976544 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976544>). PMID 21076656 (<https://pubmed.ncbi.nlm.nih.gov/21076656>).
90. Parachin NS, Gorwa-Grauslund MF (May 2011). "Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113934>). *Biotechnology for Biofuels*. **4** (1): 9. doi:10.1186/1754-6834-4-9 (<https://doi.org/10.1186%2F1754-6834-4-9>). PMC 3113934 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113934>). PMID 21545702 (<https://pubmed.ncbi.nlm.nih.gov/21545702>).
91. Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, et al. (April 2018). "Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5874163>). *Nature Microbiology*. **3** (4): 415–422. doi:10.1038/s41564-018-0110-1 (<https://doi.org/10.1038%2Fs41564-018-0110-1>). PMC 5874163 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5874163>). PMID 29434326 (<https://pubmed.ncbi.nlm.nih.gov/29434326>).
92. Raes J, Letunic I, Yamada T, Jensen LJ, Bork P (March 2011). "Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3094067>). *Molecular Systems Biology*. **7**: 473. doi:10.1038/msb.2011.6 (<https://doi.org/10.1038%2Fmsb.2011.6>). PMC 3094067 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3094067>). PMID 21407210 (<https://pubmed.ncbi.nlm.nih.gov/21407210>).
93. Lavery TJ, Roudnew B, Seymour J, Mitchell JG, Jeffries T (2012). Steinke D (ed.). "High nutrient transport and cycling potential revealed in the microbial metagenome of Australian sea lion (*Neophoca cinerea*) faeces" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3350522>). *PLOS ONE*. **7** (5): e36478. Bibcode:2012PLoSO...736478L (<https://ui.adsabs.harvard.edu/abs/2012PLoSO...736478L>). doi:10.1371/journal.pone.0036478 (<https://doi.org/10.1371%2Fjournal.pone.0036478>). PMC 3350522 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3350522>). PMID 22606263 (<https://pubmed.ncbi.nlm.nih.gov/22606263>).

94. "What's Swimming in the River? Just Look For DNA" (<https://www.npr.org/2013/07/24/205178477/whats-swimming-in-the-river-just-look-for-dna>). *NPR.org*. 24 July 2013. Retrieved 10 October 2014.
95. Chua, Physilia Y. S.; Crampton-Platt, Alex; Lammers, Youri; Alsos, Inger G.; Boessenkool, Sanne; Bohmann, Kristine (25 May 2021). "Metagenomics: A viable tool for reconstructing herbivore diet" (<https://doi.org/10.1111%2F1755-0998.13425>). *Molecular Ecology Resources*: 1755–0998.13425. doi:10.1111/1755-0998.13425 (<https://doi.org/10.1111%2F1755-0998.13425>). PMID 33971086 (<https://pubmed.ncbi.nlm.nih.gov/33971086>).
96. George I, Stenuit B, Agathos SN (2010). "Application of Metagenomics to Bioremediation". In Marco D (ed.). *Metagenomics: Theory, Methods and Applications*. Caister Academic Press. ISBN 978-1-904455-54-7.
97. Zimmer C (13 July 2010). "How Microbes Defend and Define Us" ([https://www.nytimes.com/2010/07/13/science/13micro.html?\\_r=2&pagewanted=all](https://www.nytimes.com/2010/07/13/science/13micro.html?_r=2&pagewanted=all)). *New York Times*. Retrieved 29 December 2011.
98. Nelson KE and White BA (2010). "Metagenomics and Its Applications to the Study of the Human Microbiome". *Metagenomics: Theory, Methods and Applications*. Caister Academic Press. ISBN 978-1-904455-54-7.
99. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. (March 2010). "A human gut microbial gene catalogue established by metagenomic sequencing" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3779803>). *Nature*. **464** (7285): 59–65. Bibcode:2010Natur.464...59. (<https://ui.adsabs.harvard.edu/abs/2010Natur.464...59>). doi:10.1038/nature08821 (<https://doi.org/10.1038%2Fnature08821>). PMC 3779803 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3779803>). PMID 20203603 (<https://pubmed.ncbi.nlm.nih.gov/20203603>).
100. Abubucker, Sahar; Segata, Nicola; Goll, Johannes; Schubert, Alyxandria M.; Izard, Jacques; Cantarel, Brandi L.; Rodriguez-Mueller, Beltran; Zucker, Jeremy; Thiagarajan, Mathangi; Henrissat, Bernard; White, Owen; Kelley, Scott T.; Methé, Barbara; Schloss, Patrick D.; Gevers, Dirk; Mitreva, Makedonka; Huttenhower, Curtis (2012). "PLOS Computational Biology: Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3374609>). *PLOS Computational Biology*. **8** (6): e1002358. Bibcode:2012PLSCB...8E2358A (<https://ui.adsabs.harvard.edu/abs/2012PLSCB...8E2358A>). doi:10.1371/journal.pcbi.1002358 (<https://doi.org/10.1371%2Fjournal.pcbi.1002358>). PMC 3374609 (<https://www.ncbi.nlm.nih.gov/pmc/article/s/PMC3374609>). PMID 22719234 (<https://pubmed.ncbi.nlm.nih.gov/22719234>).

---

## External links

- Focus on Metagenomics (<http://www.nature.com/nrmicro/focus/metagenomics/index.html>) at *Nature Reviews Microbiology* journal website
- The "Critical Assessment of Metagenome Interpretation" (CAMI) initiative (<http://cami-challenge.org>) to evaluate methods in metagenomics

---

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Metagenomics&oldid=1035047474>"

---

This page was last edited on 23 July 2021, at 09:56 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.